

|||||
52 GTGGCATAGAGATTTCAGACGATATAGATGCTCACTATGAT 97
seq_name: gb_pl2:AF161883

seq_documentation_block: 2182 bp mRNA PLN 10-SEP-1999
LOCUS AF161883
DEFINITION Macadamia integrifolia vicillin precursor (AMP2) mRNA, AMP2-1
allele, complete cds.
ACCESSION AF161883
VERSION AF161883.1 GI:5852871
KEYWORDS
SOURCE
ORGANISM
Macadamia integrifolia.
Macadamia integrifolia.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Proteaceae; Macadamia.
REFERENCE
1 (bases 1 to 2182)
Marcus,J.P., Goulter,K.C., Green,J.L. and Manners,J.M.
A family of antimicrobial peptides is produced by processing of a
7S globulin protein in Macadamia integrifolia kernels
The Plant Journal (1999) In press
2 (bases 1 to 2182)
Marcus,J.P., Goulter,K.C., Green,J.L. and Manners,J.M.
Direct Submission
Submitted (24-JUN-1999) Cooperative Research Center for Tropical
Plant Pathology, The University of Queensland, John Hines Building,
QLD 4072, Australia
LOCATION/Qualifiers
1..2182
/organism="Macadamia integrifolia"
/db_xref="taxon:60698"
/tissue_type="nut kernel"
1..2182
/gene="AMP2"
/allele="1"
12..2012
/gene="AMP2"
/note="MAMP2 precursor; antimicrobial peptide family
protein"
/codon_start=1
/product="vicillin precursor"
/protein_id="AAD54244.1"
/db_xref="GI:5852872"
/translation="MAINTSNLSCLFLSLFLSTVLSAESEFDROEYEECKRQCM
QLETSQMRRCVQCKRREEDIDMSKYDNDPQDCCQCCQRCRQSSGPRQOYC
ORCKETCEEEENRQDPOQOYEQCCQCHETEPHMQCQRCERREYERK
QQRKYEQOQDEDEKYEERKEEDNRDPOREYEDCRRCEQOEPQOYOCORRE
QQRHGRGQDMANPQRGSGRYEGEESQSNPYTFDEKSLSTRFTEGHISYLEN
YGRSKLRALKANRYLVLEANNPAFLPTHLDAILLVIGRGALMIIHIDNESYN
LECGDVARIPAGTFYLLNDRNERRLHIAKLTQISTPGQYKEFFPAGQNPBYLST
FSKEILFALNTQETKLVGQOQREGVITISQEOIRBELTRDQSESRHMRIRGGS
SRGPNVFNKRPYLSNKYGQAYEKPEYDROLQMDLSVFIANTQSGMGPFFNRS
TKVYVVASGADVEMACPHLSGRHGRGGRGGRHEEEDVHYEQYRALSKREAIYVLA
GHVYVVASGADVEMACPHLSGRHGRGGRGGRHEEEDVHYEQYRALSKREAIYVLA
GHVYVVASGADVEMACPHLSGRHGRGGRGGRHEEEDVHYEQYRALSKREAIYVLA
EELFNSODESIFFPGRQHQOSSRSTKQOQPLVSLDVGVF"

BASE COUNT 691 a 508 c 571 g 412 t
ORIGIN

alignment_scores:
Quality: 180.00 Length: 32
Ratio: 5.625 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-331-631-5_COPY_1_32 x AF161883 ..

Align seg 1/1 to: AF161883 from: 1 to: 2182

1 GlnCysMetGlnLeuGluThrSerGlyGlnMetArgCysValSerGI 17
|||||
135 CAATGCATGCAAGTTGAGACATCAGCCAGATCGCTGGGTGTGAGTCA 184

17 nCysAspLysArgpHegluGluAspIleasprTrpSerIysTyrAsp 32
|||||
185 GTGGCATAGAGATTTCAGACGATATAGATGCTCACTATGAT 230

seq_name: gb_pl2:AF161884
seq_documentation_block: 2182 bp mRNA PLN 10-SEP-1999
LOCUS AF161884
DEFINITION Macadamia integrifolia vicillin precursor (AMP2) mRNA, AMP2-2
allele, complete cds.
ACCESSION AF161884
VERSION AF161884.1 GI:5852873
KEYWORDS
SOURCE
ORGANISM
Macadamia integrifolia.
Macadamia integrifolia.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Proteaceae; Macadamia.
REFERENCE
1 (bases 1 to 2182)
Marcus,J.P., Goulter,K.C., Green,J.L. and Manners,J.M.
A family of antimicrobial peptides is produced by processing of a
7S globulin protein in Macadamia integrifolia
The Plant Journal (1999) In press
2 (bases 1 to 2182)
Marcus,J.P., Goulter,K.C., Green,J.L. and Manners,J.M.
Direct Submission
Submitted (24-JUN-1999) Cooperative Research Center for Tropical
Plant Pathology, The University of Queensland, John Hines Building,
QLD 4072, Australia
LOCATION/Qualifiers
1..2182
/organism="Macadamia integrifolia"
/db_xref="taxon:60698"
/tissue_type="nut kernel"
1..2182
/gene="AMP2"
/allele="2"
12..2012
/gene="AMP2"
/note="MAMP2 precursor; antimicrobial peptide family
member"
/codon_start=1
/product="vicillin precursor"
/protein_id="AAD54245.1"
/db_xref="GI:5852874"
/translation="MAINTSNLSCLFLSLFLSTVLSAESEFDROEYEECKRQCM
QLETSQMRRCVQCKRREEDIDMSKYDNDPQDCCQCCQRCRQSSGPRQOYC
ORCKETCEEEENRQDPOQOYEQCCQCHETEPHMQCQRCERREYERK
QQRKYEQOQDEDEKYEERKEEDNRDPOREYEDCRRCEQOEPQOYOCORRE
QQRHGRGQDMANPQRGSGRYEGEESQSNPYTFDEKSLSTRFTEGHISYLEN
YGRSKLRALKANRYLVLEANNPAFLPTHLDAILLVIGRGALMIIHIDNESYN
LECGDVARIPAGTFYLLNDRNERRLHIAKLTQISTPGQYKEFFPAGQNPBYLST
FSKEILFALNTQETKLVGQOQREGVITISQEOIRBELTRDQSESRHMRIRGGS
SRGPNVFNKRPYLSNKYGQAYEKPEYDROLQMDLSVFIANTQSGMGPFFNRS
TKVYVVASGADVEMACPHLSGRHGRGGRGGRHEEEDVHYEQYRALSKREAIYVLA
GHVYVVASGADVEMACPHLSGRHGRGGRGGRHEEEDVHYEQYRALSKREAIYVLA
EELFNSODESIFFPGRQHQOSSRSTKQOQPLVSLDVGVF"

BASE COUNT 684 a 510 c 572 g 416 t
ORIGIN

alignment_scores:
Quality: 180.00 Length: 32
Ratio: 5.625 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-331-631-5_COPY_1_32 x AF161884 ..

Align seg 1/1 to: AF161884 from: 1 to: 2182

1 GlnCysMetGlnLeuGluThrSerGlyGlnMetArgCysValSerGI 17

```
|||||
135 CAATCATGACGATTGGAGACATCAGCCAGATGCTCGGTGTGAGTCA 184
17 ncysaplysrpghgluaspilaspripserlystyrasp 32
185 GTCCATAGAGAGATTGAGAGAGATATGATTTGCTAAGATGAT 230

seq_name: gb_com:BTRNAT2

seq documentation block:
LOCUS BTRNAT2 2053 bp mRNA NAM 07-JUN-1995
DEFINITION B.taurus mRNA for thrombospondin (partial) 2055 bp.
ACCESSION X87619
VERSION X87619.1 GI:860884
KEYWORDS thrombospondin; TSP-1 gene.
SOURCE Bos taurus.
ORGANISM Bos taurus.
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Artiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 2053)
AUTHORS Zafar,R.S., Moll,Y.D., Womack,J.F. and Walz,D.A.
TITLE Cloning and sequencing of bovine thrombospondin stimulatory effect
of TGF-beta
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2053)
AUTHORS Zafar,R.S.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-1995) R.S. Zafar, Wayne State University, Dept of
Physiology, 540 E. Canfield, Detroit MI 48202, CANADA
FEATURES
Source
1..2053
/organism="Bos taurus"
/db_xref="taxon:9913"
/tissue_type="aortic endochelium"
/clone_lib="lambda gtl0"
/clone="pGEM-3Z(-)"
1..6
/note="EcoI cloning site"
misc_feature
180..233
/gene="TSP-1"
sig_peptide
180..233
/gene="TSP-1"
gene
180..233
/gene="TSP-1"
CDS
180..>233
/gene="TSP-1"
/function="adhesive glycoprotein"
/codon_start=1
/product="thrombospondin"
/protein_id="CAA60951.1"
/db_xref="GI:860885"
/db_xref="SPTREMBL:Q28179"
/translation="MGLAWGLGVLLHMGCS"
BASE COUNT 453 a 625 c 593 g 382 t
ORIGIN

alignment_scores:
Quality: 64.50 Length: 26
Ratio: 3.395 Gaps: 1
Percent Similarity: 73.077 Percent Identity: 46.154

alignment_block:
US-09-331-631-5_COPY_1_32 x BTRNAT2
Align seg 1/1 to: BTRNAT2 from: 1 to: 2053
7 ThSerGlyGlnMetArgArgCys...ValSerGlnCysAspLysArgph 22
:::||||| ||| ||| ||| :::::::::::|||||
1424 TCCCTGTGCGACGACGCGACCTGCCACATCCAGGAGTGTGACAAGAGATT 1473
TCTTGGTGTGCGACGACGCGACCTGCCACATCCAGGAGTGTGACAAGAGATT 1473
22 egluaspilaspripserlystyr 31
|:::::::::|
1474 TAAACAGAGATGCGGCTGGAGCCACTGG 1501
```

```
seq_name: gb_ro:MUSTSP1A

seq documentation block:
LOCUS MUSTSP1A 4339 bp mRNA ROD 09-NOV-1994
DEFINITION Mouse thrombospondin 1 mRNA, complete cds.
ACCESSION M87276
VERSION M87276.1 GI:202196
KEYWORDS thrombospondin 1.
SOURCE Mus musculus (strain CD-1) embryo kidney cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4339)
AUTHORS Laherty,C.D., O'Rourke,K., Wolf,F.W., Katz,R., Seldin,M.F. and
Dixit,V.M.
TITLE Characterization of mouse thrombospondin 2 sequence and expression
during cell growth and development
JOURNAL J. Biol. Chem. 267 (5), 3274-3281 (1992)
MEDLINE 92147683
FEATURES
Source
1..4339
/organism="Mus musculus"
/strain="CD-1"
/db_xref="taxon:10090"
/dev_stage="embryo"
/tissue_type="kidney"
212..3724
/codon_start=1
/product="thrombospondin 1"
/protein_id="AAA53063.1"
/db_xref="GI:567240"
/translation="MELRGLGLVFLHMGCSNRIPESGGDNGVDFELIGARGRP
GRRLVKGDDISSPAFRINAMNIPAVPDDKRODLIDAYMARKGFELASIROMKTRG
TLAVERNNGQISFVSNGKAGTDLISLPKQVAYSEALALAGOKSTLFLV
QEDRAQLYTDCKMSAEIDVPIQSTIFTRDLASVARKLVAGVNDVQVQVNRFLV
FETTPEDILRNKGCSSSTNVLTFTDNNVNGSSPAIRNVIQHKTRDQAICGSCDE
LSMWELKGLRTIVTTLQDSIRKVTENRELVSLEKRPILCFHNGVOYKNNIEWTVS
SCTHECONSVTICKVSCPIIMPCCNAIVPGECCPRCPSPSADGSPMSSEVTS
ATCGNGIOGRGSCDSLNRGESSVQVRYRTHIOGRDRFRDGGSHMSPSGCVT
CGDYVTRIRILRCLSPSPOMNKRPGCEGARETRKACKKDKCPTINGSGNPGSPMDICVTC
GGGVDRSRILCNPTPQFGKDCGVGVLENVCNKKDQCPIDGUSNPGCFAGKCTSTP
DGSWKGCAGPQYSGNGIQCKDVDECKVPDPACFNHNEHRCKNTPDQYNGILPCPRF
TGSQPFGRGVEHMANKQVCKRPNCPTGTGHDCKNNAKCNVILGHSYDPMYCECKPGY
ANGGIIICEPDLDLQGMENLVYVANAATYHCKKDKCPMLPNSGQDYDKDGIQDACC
DDDNKRTEDDNDGCPENHYNPAQYDVRDQDRCDCNCPYNNPDQADDTKNGEGGACA
VDIDDGILNERNDQYVYNNDRPTDNDGVGDQDNCPLHNPDQLDSDSDILIDTC
DNQDIDEDGHQNNLDNCPYVPMANQADHDKDGKDACDHDNDGIDPDDNCKLVP
NPQKDSGDGKRGKADKDFHDVNPDIIDICPENFDISEDFRFRQMIPLDPKTSQ
NDPMNVVHRQCKELVQTVNCDPGLAVGDEFENAVDFSGTFINTERDDVAGLVFGYO
SSRFYVVMKQVQSYMDTNPTRAQYSGSVKVNSTTGPGEHLRALMHTGNTPG
QVRLTMHPRHIGKMDPTAYRRRLSHRKRTGIYRVMEGKKIMADSGPIYDKTYAGC
RLGLVFQDENVFEFSDMKYECRDS"
BASE COUNT 1109 a 1201 c 1144 g 885 t
ORIGIN

alignment_scores:
Quality: 64.50 Length: 26
Ratio: 3.395 Gaps: 1
Percent Similarity: 73.077 Percent Identity: 46.154

alignment_block:
US-09-331-631-5_COPY_1_32 x MUSTSP1A
Align seg 1/1 to: MUSTSP1A from: 1 to: 4339
7 ThSerGlyGlnMetArgArgCys...ValSerGlnCysAspLysArgph 22
:::||||| ||| ||| ||| :::::::::::|||||
1457 TCTTGGTGTGCGACGACGACCTGCCACATCCAGGAGTGTGACAAAAGATT 1506
TCTTGGTGTGCGACGACGACCTGCCACATCCAGGAGTGTGACAAAAGATT 1506
22 egluaspilaspripserlystyr 31
|:::::::::|
1507 TAAACAGAGATGCTGGAGTCACTGG 1534
```

seq_name: gb_pr2:HSTHRMR

seq_documentation_block:

LOCUS HSTHRMR 4434 bp mRNA

PRI 21-MAR-1995

DEFINITION Human mRNA for thrombospondin.

ACCESSION X04665.1 GI:37137

KEYWORDS glycoprotein; signal peptide; thrombospondin.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 4434)

Lawler, J. and Hynes, R.O.

The structure of human thrombospondin, an adhesive glycoprotein

with multiple calcium-binding sites and homologues with several

different proteins

J. Cell Biol. 103 (5), 1635-1648 (1986)

87057617

JOURNAL

COMMENT

FEATURES

SOURCE

Three types of repeating amino acid sequence are present in
thrombospondin. The first is 57 amino acids long and shows homology
with circumsporozoite protein from Plasmodium falciparum. The
second is 50-60 amino acids long and shows homology with epidermal
growth factor precursor. The third occurs as a continuous eightfold
repeat of a 38-residue sequence; structural homology with
parvalbumin and calmodulin indicates that these repeats constitute
the multiple calcium-binding sites of thrombospondin. Data kindly
reviewed (15-SEP-1987) by Lawler J.

location/Qualifiers

1..4434

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_type="umbilical vein endothelial cells"

76..3588

/note="precursor polypeptide (AA -18 to 1152)"

/codon_start=1

/protein_id="CAA28370.1"

/db_xref="GI:37138"

/translation="MGLAMGLVLLIHAAGSGNRIPESGDNVDFIFELTGAAARKS

GRILVKGDPSSPAFRIDANLIPVPDPRKQDLYDVAIRTEKGLILLASLQMKRTKG

TLALERKDHSGVSVSNKAGLTLVQKQHVSVSEALATGQKSTLLE

QEDRAQLIDICEKEMNELDVPVQSVFTRDASLARLHAGVNDVFOALGVNREV

FGTTPEDILRNKGCSSSTVLLTDNNVNSSPAIRNTYGHKTQDLQALGISCDE

LSMWLELRGLRTVTTLDISIRKYTEENKELAREPPLCYNGNVOYRNEMETVD

SCTECHONSVTICKVSCPIPCSNATVPECCPCPMPSDADDCGSPSWSCST

TSCNGITQQRSCDSLNRCESSVQRTTIOEDCKRFQDDGSRHSPWSCST

CGDEVITIRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

misc_feature 3226..3234
/note="pot. N-glycosylation site"
misc_feature 3274..3282
/note="pot. N-glycosylation site"
BASE COUNT 1139 a 1185 c 1177 g 933 t
ORIGIN

alignment_scores:

Quality: 64.50 Length: 26

Ratio: 3.395 Gaps: 1

Percent Similarity: 73.077 Percent Identity: 46.154

alignment_block:

US-09-331-631-5_COPY_1_32 x HSTHRMR

Align seg 1/1 to: HSTHRMR from: 1 to: 4434

7 ThrSerGlyIleMetArgArgCys...ValSerGlnCysAspLysArgph 22

1321 TCCTCGGTCACAGACGACCTGCACATTCAGAGTGTGACACAGAGATT 1370

22 egluGluaspIleaspTrpSerLysTyr 31

1371 TAACAGGATGGTGGCTGAGCCACTGG 1398

seq_name: gb_cm:AB005287

seq_documentation_block:

LOCUS AB005287 5269 bp mRNA

DEFINITION Bos taurus mRNA for thrombospondin 1, complete cds.

ACCESSION AB005287

VERSION AB005287.2 GI:4630795

KEYWORDS thrombospondin 1.

SOURCE Bos taurus

ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;

Bovinae; Bos.

1 (sites)

Ueno, A., Yamashita, K., Negata, T., Tsurumi, C., Miwa, Y., Kitamura, S.

and Inoue, H.

cDNA cloning of bovine thrombospondin 1 and its expression in

odontoblasts and preodontin

Biochim. Biophys. Acta 1382 (1), 17-22 (1998)

98173773

2 (bases 1 to 5269)

Ueno, A.

Direct Submission

Submitted (30-JUN-1997) to the DDBJ/EMBL/Genbank databases.

Akemihi Ueno, The University of Tokushima, Department of

Biochemistry, School of Dentistry, Kuramoto-cho 3-18-15, Tokushima,

Tokushima 770, Japan (E-mail: akemihi@dent.tokushima-u.ac.jp,

Tel: +81-886-33-7326, Fax: +81-886-31-4215)

On Apr 21, 1999 this sequence version replaced gi:2244706.

location/Qualifiers

1..5269

/organism="Bos taurus"

/strain="Holstein"

/db_xref="taxon:9913"

/cell_type="odontoblast"

/clone="BPTSI-28"

/dev_stage="two weeks after birth"

/sex="male"

/tissue_type="anterior tooth"

/codon_start=1

/product="thrombospondin 1"

/protein_id="BA21115.1"

/db_xref="GI:2244706"

/translation="MGLAMGLVLLIHAAGSGNRIPESGDNVDFIFELTGAAARKS

GRILVKGDPSSPAFRIDANLIPVPDPRKQDLYDVAIRTEKGLILLASLQMKRTKG

TLALERKDHSGVSVSNKAGLTLVQKQHVSVSEALATGQKSTLLE

QEDRAQLIDICEKEMNELDVPVQSVFTRDASLARLHAGVNDVFOALGVNREV

FGTTPEDILRNKGCSSSTVLLTDNNVNSSPAIRNTYGHKTQDLQALGISCDE

LSMWLELRGLRTVTTLDISIRKYTEENKELAREPPLCYNGNVOYRNEMETVD

SCTECHONSVTICKVSCPIPCSNATVPECCPCPMPSDADDCGSPSWSCST

TSCNGITQQRSCDSLNRCESSVQRTTIOEDCKRFQDDGSRHSPWSCST

CGDEVITIRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC

DGQVQKSRILCNSTPQPGKDCVQDVTEENICKKDCPLINGGWPSPWDICSTVC


```

Align seg 1/1 to: HSTS from: 1 to: 5722
7 ThisSerGlyInMetarGargCys...ValSerGlnCysAspLysArgPh 22
:::||||| ||| ||| ||| :::::::::::|||||
1357 TCCTCGTTCACAGACGACCTGCGACATTCAGAGAGTGTGACAAAGATT 1406
22 eGIUGLAspIleAspTrpSerLysTyr 31
:::||||| ||| ||| ::
1407 TAAACAGAGATGTGCTGAGACCACTGG 1434

seq_name: gb_ov:XELSTAGE

seq_documentation_block:
LOCUS XELSTAGE 3522 bp mRNA VRT 12-JUN-1993
DEFINITION Xenopus laevis mRNA sequence.
ACCESSION L04278
VERSION L04278.1 GI:295541
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis stage 45 cDNA to mRNA.
Xenopus laevis
Eukaryota; Metazoa; Chordata; Vertebrata; Amphibia;
Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodidae;
Xenopus.
REFERENCE
1 (bases 1 to 3522)
Urry,L.A., Ramos,J., Duquette,M., Desimone,D.W. and Lawler,J.
Cloning, characterization and expression of thrombospondin-1 in
Xenopus laevis embryos
JOURNAL
Unpublished (1992)
FEATURES
Location/Qualifiers
1..3522
/organism="Xenopus laevis"
/db_xref="taxon:835"
/dev_stage="stage 45"
ORIGIN
BASE COUNT 1031 a 786 c 884 g 821 t
ORIGIN

alignment_scores:
Quality: 62.50 Length: 26
Ratio: 3.289 Gaps: 1
Percent Similarity: 73.077 Percent Identity: 46.154

alignment_block:
US-09-331-631-5_COPY_1_32 x XELSTAGE ..
Align seg 1/1 to: XELSTAGE from: 1 to: 3522
7 ThisSerGlyInMetarGargCys...ValSerGlnCysAspLysArgPh 22
:::||||| ||| ||| ||| :::::::::::|||||
1255 TCCTCGTTCACAGACGACCTGCGACATTCAGAGAGTGTGACAAAGATT 1304
22 eGIUGLAspIleAspTrpSerLysTyr 31
:::||||| ||| ||| ::
1305 TAAACAGACGAGTGTGAGCCATTGG 1332

seq_name: gb_hcgl14:AC020788

seq_documentation_block:
LOCUS AC020788 197071 bp DNA HTG 19-FEB-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-162J10, *** SEQUENCING IN
PROGRESS ***, 113 unordered pieces.
ACCESSION AC020788
VERSION AC020788.2 GI:7007946
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 197071)
Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodoty,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunec,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
Hollway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondewski,N., Kong,X., Kovar,C., Leal,B., Li,Z.,
Lichter,R., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Oulles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugang,R.,
Taber,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
Worley,K., Wren,J., Wrenford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 197071)
Worley,K.C.
Direct Submission
Submitted (10-JUN-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 19, 2000 this sequence version replaced gi:5686414.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMXE
Center clone name: RP11-162J10
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; version 0.980611
Assembly program: Phrap; version 0.980611
Consensus quality: 103326 bases at least Q40
Consensus quality: 135876 bases at least Q30
Consensus quality: 151983 bases at least Q20
Estimated insert size: 165698; sum-of-coverage estimation
Quality coverage: 1.2x in Q20 bases; sum-of-coverage estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 113 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
785: contig of 785 bp in length
786 805: gap of unknown length
786 1571: contig of 766 bp in length
1572 1591: gap of unknown length
1592 2373: contig of 782 bp in length
2374 2393: gap of unknown length
2394 3161: contig of 768 bp in length
3162 3181: gap of unknown length
3182 3945: contig of 764 bp in length
3946 3965: gap of unknown length
3966 4779: contig of 794 bp in length
4780 4779: gap of unknown length
4780 5649: contig of 870 bp in length
5650 5669: gap of unknown length
5670 6452: contig of 783 bp in length
6453 6472: gap of unknown length
6473 7253: contig of 781 bp in length
7254 7273: gap of unknown length
7274 8074: contig of 801 bp in length
8075 8094: gap of unknown length
8095 8863: contig of 769 bp in length
8864 9669: contig of 786 bp in length

```

```
* 9670 9689: gap of unknown length
* 9690 10484: contig of 795 bp in length
* 10485 10504: gap of unknown length
* 10505 11291: contig of 787 bp in length
* 11292 11311: gap of unknown length
* 11312 12093: contig of 782 bp in length
* 12094 12113: gap of unknown length
* 12114 13121: contig of 1008 bp in length
* 13122 13141: gap of unknown length
* 13142 14243: contig of 1102 bp in length
* 14244 14263: gap of unknown length
* 14264 15458: contig of 1195 bp in length
* 15459 15479: gap of unknown length
* 15479 16343: contig of 864 bp in length
* 16343 16363: gap of unknown length
* 16363 17521: contig of 1159 bp in length
* 17522 17541: gap of unknown length
* 17542 18227: contig of 686 bp in length
* 18228 18248: gap of unknown length
* 18248 19001: contig of 754 bp in length
* 19002 20081: gap of unknown length
* 20082 20101: contig of 1060 bp in length
* 20102 21503: gap of unknown length
* 21506 21525: gap of unknown length
* 21526 22240: contig of 715 bp in length
* 22241 22260: gap of unknown length
* 22261 23639: contig of 1379 bp in length
* 23640 23659: gap of unknown length
* 23660 24376: gap of unknown length
* 24377 24396: gap of unknown length
* 24397 25517: contig of 1121 bp in length
* 25518 25537: gap of unknown length
* 25538 26325: contig of 788 bp in length
* 26326 26345: gap of unknown length
* 26346 27441: contig of 1096 bp in length
* 27442 27461: gap of unknown length
* 27462 28609: contig of 1148 bp in length
* 28610 28629: gap of unknown length
* 28630 28912: contig of 1283 bp in length
* 28913 28932: gap of unknown length
* 28933 30884: contig of 952 bp in length
* 30885 30904: gap of unknown length
* 30905 31688: contig of 784 bp in length
* 31689 31708: gap of unknown length
* 31709 32733: contig of 1025 bp in length
* 32734 32753: gap of unknown length
* 32754 33401: contig of 648 bp in length
* 33402 33421: gap of unknown length
* 33422 34317: contig of 896 bp in length
* 34318 34337: gap of unknown length
* 34338 35695: contig of 1358 bp in length
* 34339 35715: gap of unknown length
* 35696 35716: gap of unknown length
* 35717 36961: contig of 1246 bp in length
* 36962 36981: gap of unknown length
* 36982 37808: contig of 827 bp in length
* 37809 37828: gap of unknown length
* 37829 38513: contig of 685 bp in length
* 38514 38533: gap of unknown length
* 38534 39273: contig of 740 bp in length
* 39274 39293: gap of unknown length
* 39294 40153: contig of 860 bp in length
* 40154 40173: gap of unknown length
* 40174 41306: gap of unknown length
* 41307 41326: gap of unknown length
* 41327 42120: contig of 794 bp in length
* 42121 43356: contig of 1216 bp in length
* 43357 43376: gap of unknown length
* 43377 44189: contig of 813 bp in length
* 44190 44209: gap of unknown length
* 44210 44837: contig of 628 bp in length
* 44838 44857: gap of unknown length
```

```
* 44858 45536: contig of 679 bp in length
* 45537 45556: gap of unknown length
* 45557 46310: contig of 754 bp in length
* 46311 46330: gap of unknown length
* 46331 47381: contig of 1051 bp in length
* 47382 47401: gap of unknown length
* 47402 49139: contig of 1738 bp in length
* 49140 49160: gap of unknown length
* 49160 50915: contig of 1756 bp in length
* 50916 50935: gap of unknown length
* 50936 52088: contig of 1153 bp in length
* 52089 52108: gap of unknown length
* 52109 53346: contig of 1218 bp in length
* 53347 53366: gap of unknown length
* 53367 54978: contig of 1632 bp in length
* 54979 54998: gap of unknown length
* 54999 55817: contig of 819 bp in length
* 55818 55837: gap of unknown length
* 55838 56692: contig of 855 bp in length
* 56693 56712: gap of unknown length
* 56713 57776: contig of 1064 bp in length
* 57777 57796: gap of unknown length
* 57797 59164: contig of 1368 bp in length
* 59165 59184: gap of unknown length
* 59185 60494: contig of 1310 bp in length
* 60495 60514: gap of unknown length
* 60515 61716: contig of 1202 bp in length
* 61717 61736: gap of unknown length
* 61737 62752: contig of 1016 bp in length
* 62753 62772: gap of unknown length
* 62773 64074: contig of 1302 bp in length
* 64075 64094: gap of unknown length
* 64095 65349: contig of 1255 bp in length
* 65350 65369: gap of unknown length
* 65370 66810: contig of 1441 bp in length
* 66811 66830: gap of unknown length
* 66831 68613: contig of 1783 bp in length
* 68614 70193: gap of unknown length
* 68634 70193: contig of 1560 bp in length
```

```
alignment_scores:
  Quality: 62.50      Length: 31
  Ratio: 2.976       Gaps: 1
Percent Similarity: 67.742 Percent Identity: 41.935
```

alignment_block:

US-09-331-631-5_COPY_1_32 x AC020788/rev ..

Align seg 1/1 to reverse of: AC020788 from: 1 to: 197071

```
1 Gtncysmctglnleugluthrsrclgylmctargacysvalsercl 17
|||||  ||||||  |||  |||  |||  |||  |||  |||
30657 CAGTGT...CAGCTGTGTGGAAGTATCAGAAAGCTCTCAGACGCA 30611
```

```
17 ncysaplysarqphcgluaslplleasprserlystyr 31
|||||  :::::::::::  |||  :::::
```

30610 ATGTCACAGCTCATGAGAGTCTGAAGTGAAGTGCCTCCG 30568

seq_name: gb_ba2:AF188287

seq_documentation_block:

LOCUS AF188287 42603 bp DNA BCT 12-JAN-2000

DEFINITION Stigmatella aurantiaca myxochiazol gene cluster, complete sequence.

ACCESSION AF188287

VERSION AF188287.1 GI:6635393

KEYWORDS Stigmatella aurantiaca.

SOURCE Stigmatella aurantiaca

ORGANISM Bacteria: Proteobacteria: delta subdivision: Myxobacteria:

Myxococcales: Cystobacterineae: Cystobacteraceae: Stigmatella.

1 (bases 1 to 42603)

Silakowski,B., Schaefer,H.U., Ehret,H., Kunze,B., Weinig,S.,

Nordstiek,G., Brandt,P., Bloeker,H., Hoeffle,G., Beyer,S. and

TITLE
Mueller, R.
New lessons for combinatorial biosynthesis from myxobacteria. The
myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca*
DM4/3-1

JOURNAL
J. Biol. Chem. 274 (52), 37391-37399 (1999)

MEDLINE
20069734

REFERENCE
2 (bases 1 to 42603)

AUTHORS
Silakowski, B., Schaefer, H.U., Ehret, H., Kunze, B., Weinig, S.,
Nordstedt, G., Brandt, P., Bloeker, H., Hoeffle, G., Beyer, S. and
Mueller, R.

TITLE
Direct Submission

JOURNAL
Submitted (20-SEP-1999) NBI/MX, GBF, Mascheroder Weg 1,
Braunschweig 38124, Germany

FEATURES
Location/Qualifiers
1. 42603
/organism="Stigmatella aurantiaca"
/strain="DM4/3-1"
/db_xref="taxon:41"
1. 834
/gene="mtaa"
1. 834
/gene="mtaa"
/note="pent transferase"
/codon_start=1
/transl_table=11
/product="Mtaa"
/protein_id="AAF19809.1"
/db_xref="GI:6635394"
/translation="MPTSSPALDLKLPDEYHVIVPEPRTEPGILFSSYALIDPG
ENDKORFFYFERRLQYLVSHALVRLTISRIVAPPEANSFPAENOYGREINDEEPEW
LRFNSHTDGMALCAVARDVDGADVEDERGETEVLADSFPAENYLSLALPVSG
QERREFDYWTLEAVIKARGMGLSEPLDQFAFESQGLSTRISFPRLVDESOQVCF
RRSPQRHAAALAVRRSEAPLTVRFQRTVPILDADPAEYLRSERIQPLRMPCVGC
G"
1388. 13399
/gene="mtab"
/note="PKS"
1388. 13399
/gene="mtab"
/note="PKS: Mtab loading domain and module 1: polyketide
synthase domains: ACP, KS, AT, DH, ER, KR, ACP; Mtab
module 2: polyketide synthase domains: KS, AT, DH, KR,
ACP"
/codon_start=1
/transl_table=11
/product="Mtab"
/protein_id="AAF19810.1"
/db_xref="GI:6635395"
/translation="MSTERGVGVGRSTGQLRDMLVQVAVSGVLEAHTVRITPEFS
LGVSVFATRLIAELARMLERPLSPTLVWEHPTIDYFAHLSGAEPPEAMHAPGEA
AVDGAPEIATVIGLACRFPRAPAKAFWYLGQRDCVTEPSNRMDASLYYPAEPT
GIVTRYGAFDSVDRDFPGISPREAGOLDPOORLMLSEWLEALGAGIAPRILR
GRTGVETGMWRVDADLOTINAAPTLHTCTGACGIIANRISYALGEGSLAIDT
ACSSSLVAIHEACRLSRGESSLADGGVQLNSPRTFVAVSAGALADGRCRTFDA
RASGVYRGGGVLVLRPLSRALADDAIYCLIRGSVANNDASNGLAPNKAQERL
LBDAYAVAGMDPAQVHYVELHGTGTPGLGDPITAGALVAGIARAEARLELHVSQYTN
IGHLEAAAGIASYIKVALSIRHQULPASLPEOPNRIPEDDLRLVQQLAPWPAEK
DPAVAGISAFGGGTGCHVYVEEHREGSRTAPVSTPAOSPRHKPTKQALGASA
WMKMGRELGAEPVPARAREEDDANRPLTGMSIVPALLAPVRSREDEGEQDLPLV
ANQISLALMRSMGIEPSATVGSVGENAAOVAGLITILEFALATVHTIQVKGAG
GVGMAYVDSGAENAVRIAPYGGRIISASENGRSTGLGEAAALODFTASTAGTIF
ARLVEIDFVAHLPQVETALPLFLOSCPVLRPSGRPWPSTVSGAFVDEGTIGAEYWA
NSREVPFATAMRTLLEGYDVLVDVNHPIHESVLEENRAGTASAEFLSMRGE
PSEVEGHRALTOLOSAGCPVWRERVPVPGAGSAPVRELFVAASARTBOALBEADR
IAEQLRDAASSTLDTCTAGARSHHEHRLALVATRAALEGLAAFRGRTIRPSS
QGVRPEGAPRLGFEVSQGSFPLGSSVOLFEESPFRAERSEKLEFPAQGSALIE
ALHAERSEHPRTTEIOPLEPLFOVALAWESWGIRPDAYGHSVGSISAAHAGV
LSLEALSLVSVSRMLGRAGAGGMAVGAEPARIEPLGKCDLTLAAANGENLV
VYSGALAAALASRLDELTOGLRVRVNVYARFSAQMAIYLGDLPLVLEAKPKPRL
PIYTVHGRFLASASDGTGDMWRANVERILFAPAYAMRABDGHAFVELSPRPVLOP
LSRLTIDASPASGASALFTRRDBAPRAALGLNGLITYLGHESPMSKGVPPDLGRYEL
PSTPMORQRCWFDAAQPLAEVHPULGRRVAPPGQIVFEATFNSDRQLQLSHRLLY
GTVVAVAGASMLGMILVAARQIWGEAFCALQDLVLFLOPLVLDGTEDRLVQVSTATDGA

RILSWAGOGEGTSSWHIHAATGRLEGAEETENGREVLSEVENARCPPLASASFEYE
GIARQGLGOFRLFEALMOGDRAFRLEVEETDCAEDYRFRHGLIDSCFOYIGA
VAGFRTIGLAEFAPVIGVGRFALRGARPRFCHVLRSGSGKANASSVADLILDESG
LVYAVESGLVTRRAPREYMLRDYTAAPLVYRTENQPMOMSGASQAGEALILADRSS
GLARLAEOULSRGSRSEVLAENAGSMULRTEIHDISGLACAGVDRKUDDESSS
DAVOGAVRVQDLVTRRAEARLVLYVTRGVQVGSQDGSIAFEDAPLMWIGMNTALE
HPELRCORIDLDPSGPADAQALAEALVLAEPREDEAVAFREGRRLVARVTRKLPVAV
ALPLEGDAYVIVGLGIGLELAWMYGRARHLYVLGRSAPASHAQETLALADGAG
ATISVALGNAVEREARLAVGLDIRAMPALGVNHAALVLDGLVLRSEERLAVLA
PKIOGAMNLHRETVGRDLRFVLTSSVSYSVGTGTCISDVALLELNSDPDGINAYSG
TALSINMPADVGMARLPPEDACKSNSEFGLIAPASGIALEERLISSTRAOVYA
GASEAEAGDFAEPFDALSRKPEASRASGVGLVGSPELRLRLSLLOVQATV
LGLPEKSIDPRRLQDELGLDSLALQNRSLASLALPALTLEPHTLEKLEFL
ARDLTPASEASPAKSRPVKLPARABEAEALVAVGMAKRLPGVSRPEEPRKLLANG
VNAREVEKORMSLDTWYDPDEPFGKSHSGGGLDEVDPEDFASFCISAPEARMS
POORLILEGMEALASAGIAVDOLKSGTGTGTCISDVALLELNSDPDGINAYSG
SGCVISVAGRIAYVLCGEGTIAVDTCSSIVAAHILACOSLRGECEDIALVGCNLS
LISPRMSYFSTRKALSPDGACAKFADSSANGYVREGAGVVLKRLSEALADGDTLS
VMRGTAVQDGRSNGLTAPSRQAQERVIESALROGAVSPLDVSYEARGTSLGDP
EVOALAAVIGKDRPSAPLIGISVYTNHGLEAGAVSLIKAILALRHRLPSRLH
KTPSPHIMDELPVEVISEHRPMVPEGRRLIGVSRFGSTNAHVILEEAPGPR
SKAPSIPEPRPELVISARAPOLQEAARVRESAOLSGRAGDAAMHDVCSANCRRSH
HEHRLAVYARSTPEAEALAEFLRLAPVDVQAGLALGAPRVYFLETGGCGOVRGT
GRALKEEPRFERSALQADDEVKLGAGSIIEELQREPRTDQAIISQAPHALQASAL
IALRSMGIEPDVAGSSIGVAAVYAGVLTLEAMRTIVARSLIHAVAGKGMAL
VELSEEAETTEAAVYAOQLSYAALINSQSTLSCGAMDELLGALTARGVSCRIGL
DYASPAVEPEYIDELKOLOGLRPOPAITIFSVTRGLAEAGFNEDYARNLRP
VRLAPVLTTLAREDTLFLVELGPNVAVMAPVOCCQVQAKSATLITLSTRGSEERHTL
RSVNGDLEIACVTPANNMLPFEQGSIVSLSPYPMOKEXYVNASPSALQIPASQCV
LHGRVNRSPLESLETDLSYKMFPLEHLLHGLVAGAHVALALSAIRELWGLA
PVRLADIVPEPLVLTKEGVENPVAVVGRGGRALIGSPQSEPEPMVHAEWVLA
PASQSEPAQVLAELERARCPNGLESPVYKADAEIGISGRFOWEOLMOGGEA
IGRMAEASADQHPGAPLPHGLIDSIIFVLKADGIGGDCVPYIPIDIEELCLAP
RGVLAHVYKRLREGSROGETLADRLPEESGKPAWIDGLIVKRAPEARORHAEOD
LSEKIFQVROMSMPLARVPRGRPLVADGPRAGELEALRLKREGOCHVLAPESA
VTGELQVREBALLEAVIKSVYVLSIAEADQESQGSQALBESVGAVALQAIK
LERGTPRLMLVTRGAQALLESPPRLAQAQPLWGLGLAVDEHPELCTLIDIDP
AADGVGCLFOELTLPDRDRIAMRGAFVRLGLEAVPEPTAPISPDCTYLTGG
LGGIGLCARMLVDGARHIALGLSGSEASAVGLAELEAGVRYVLAQADVAVER
VSKYLSAIAVRMPLRGVLAHAGLIDDLALROBORRORVLOVLOKVGALNLDIETRG
DIDPEFVYFSSITAMLTGEGCGYSAANAFLDALAFARAGGIPRHSILAMPAYCMA
VVAERBSAHLAALGIESITPPDEGVAILGILLRPAASAVNGYFQIDMTWQORHAGHA
PLLAGILKQQAAPSDSSEAPDIOELIRATPERORAAEFLRREVTRLDLPRL
ADGRLPVIRVGFDSLMSLKLKARLESTIGLSIAKALGSLDLKVALVLERITLQ
PTPVSTDOMEDFRP"
13537. 17409
/gene="mtac"
13537. 17409
/gene="mtac"
/note="peptide synthetase; peptide synthetase domains:
heterocyclisation, adenylation, PCP note: motif with weak
homology to putative NAD binding proteins from pyrococcus
species located behind PCP; this motif may be involved
in-thiazole formation from the proposed thiazoliene
intermediate of the biosynthesis of myxothiazole"
/codon_start=1
/transl_table=11
/product="mtac"
/protein_id="AAF19811.1"
/db_xref="GI:6635396"
/translation="MIELLRKEVAPDAQFEPFPLTDIOETVFGNGTDFGSGMS
CHLYIELDTRMDVARLSRAMOQLLEQVMLRSVLPFGEOVRLESVPFSPIDULR
GOAPLVETLKAELIRQOGLSTSTPGGOWPTEFLRATVLDDEGRTLRHGFEDLTDMDP
VLTILEERREYLSDRDLKQVPSFQPHVLAEEAARPTDSFQRAEYVWNRALNDAP
APALDVATEPERLARPSFRRLSGRLDAARAKRLERANACLTISGVCAAFAYELGA
WSESREFTVLRRLRLEIPEVRLDFTTNVLVDRGSEFTAVAVLRLQOALE
DLEHREFTVLMRQAMQKNGAIMPVVTSLIGHNSRVPGGGSPSEWGLVEVYAI
TQTPOMTMDVYREEGALYCHMPSDEGLPEGLGGEFAVYTLTLLSLADNLMEE
GTVVLPAGQABRETMAATYASIPEARLDGLFLAOGCGORMALIPEALRMSYGL
LGQGSIAHIAELGARPNETLVAYLEGCMQIYAYTAIHAGSAYLPIDALPEER
RLLEEGQKVVLTLSHLERLSPASRYVAVDTAGAPRRLPARASAGACITIT
SGSTRPGAVYIEHRAANTLIDINERGVPAADVATLSSGLDLSYVDLFGSLAG
AAIVLPOPEVYRDBAHMWLOVLRQOGVITMNSVPTLMEMLVQVVEARGELPELTRLVMA

LSGDWIPVLPDRIKIRLSDIRIVISLGATEASIMSIINPGRVDPAAIRSPYGRPL
 NORRYVDEKLAPDPHVPDGLYIGVGLAQGFREDDETARFENPGRGELRYG
 DLGFLPEGEIEFGKDFQVLAAGRIIEGEIEAVIQRHPSIREAVANAGERTSR
 LVAYIVPEGEADSPDEVLSPFLAALPQWNAVIVFRLHPLSSNGKDRALPP
 VLPDSRASGLIDARSGKITLSAKMLSEVRLSRISIEBGLQIGATVLLIKLSML
 EOEFGSPRTITTELLSLNASEIAIAYVERPAERASDSAEIIPAIREFDGAAYA
 QLEAPPEDEAROCYEFKRVSRHRTFSAPVPAVLAALSLCLSPKYGQKYGSGAG
 GIYGVOTYLSIATGRVGLDAGAYYDPLRHTTVOLGASGLPALDSSANMAADA
 AFSYLIGDRRVISORGERMDALIEAGLMAQLLETRAESDLGICQLELREFEL
 SKALREEHMYLHGLVGGSLAMEGAL"
 17406..27281
 /gene="mtad"
 17406..27281
 /gene="mtad"
 /note="PKS/NRPS: Mtad module 1: peptide synthetase
 domains: heterocyclisation, adenylation, PCP note: motif
 with weak homology to putative NAD binding proteins
 from Pyrococcus species located between adenylation core
 regions A9 and A10; this motif may be involved in thiazole
 formation from the proposed thiazoline intermediate of the
 biosynthesis of myxothiazol; similar to the motif found in
 Mtac, Mtad module 2: polyketide synthase domains: KS, AT,
 DH, KR, ACP"
 /codon_start=1
 /transl_table=11
 /product="Mtad"
 /db_xref="GI:6635397"
 /translation="MSTSERIVYIQQKIGILMEIGDTLRFRAPPGHITPLRLAKDK
 KEQLLELRASASRANLQTVASPPDASRAPEPLDIQNAVWVGAGADAGVAAH
 SYLEADTDLRLLESILQRLIEYHMLRMVYLPTEGVVLEKVPFTLTADFLAG
 PASOIEAHFEALRIGELSHOVLPTDRFLFVPAKSLAGRGVHNSILATLAKSVQ
 ILLEOCSTLYLNEAOLVPRKRTFEDVCVLAORADPSRAVNSIAYMNRLEKLPN
 ADPLPIVNDVGOHOFVRKAGIQOASRRFRNNAKACITRSMALCAVCEETARA
 YSRNRLTLNLTLPENRPLIEDVRYIGDTSGILLESKSPESFAARRKROGDT
 NDLEHSAVSVKVLREAIRLRWDAGTAMPYVFTSLISERGLSYSEGVKIVDIVSQ
 TPQWLDHOFLELDGIGYMSVDALFPAGLIDAVESYGRILHRLSSASAMALS
 ROPLDLPOLASRAEVNATERVSEQRLESFLROAEROPGAVAVIAGEVLSGKLER
 RSRARIAWLIAGCKPDOLVAIVAEKGEQVVAALIRAGAIVLPSPILPEPRVHA
 LREFAADVLTQSALESLSRMEGPPRLSVDRDEPELARKGALPPQGNNAIYVIT
 SGTGRKGYMIDHRGAVNTLLDMNEFNNGPDRVIALSLSFDSLVDLEFVLAG
 GAIVMEPGTSRDPGRWQVILEKTVITNSVPAALMDLVEFSEGLRLPLSLRLVL

alignment_scores: Quality: 62.00 Length: 25
 Ratio: 3.100 Gaps: 0
 Percent Similarity: 80.000 Percent Identity: 40.000

alignment_block:

US-09-331-631-5_COPY_1_32 x AF188287 ..

Align seg 1/1 to: AF188287 from: 1 to: 42603

5 LeuglurhSergIyGlnMetargCysValSergIInCysAspLysar 21
 |||:||||| ::|||:|||||:|||||:|
 29116 CTCCGAGACTCTCCGCGTCTCCGCAAGACATCGAATCGACACGAC 29165
 21 gPhegluGluApeIleAspTrpSer 29
 |||:| ::|||:|
 29166 CTTCGGGGCCCATGTGACTGTCT 29190

seq_name: gb_hgt10:AC019039

seq_documentation_block:

LOCUS AC019039 162168 bp DNA HTG 06-FEB-2000
 DEFINITION Homo sapiens chromosome 18 clone RP11-390D24 map 18, LOW-PASS
 SEQUENCE SAMPLING.
 ACCESSION AC019039
 VERSION AC019039.2 GI:6922690
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE ATTORNS TITLE JOURNAL REFERENCE AUTHORS

TITLE
 JOURNAL
 COMMENT

Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 162168)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 18, clone RP11-390D24
 Unpublished
 2 (bases 1 to 162168)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bede, F.,
 Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 Dearrellano, K., Dewar, K., Domiano, M., Doyle, M., Fenesfor, J.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Landers, T., Lehoczy, J., Levine, R., Lien, C., Liu, G., Locke, K.,
 Macdonald, P., Margulis, N., McEwan, P., McGunk, A., McKernan, K.,
 McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
 Pierre, N., Pisanli, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talmas, J., Tesfaye, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
 Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (30-DEC-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Feb 7, 2000 this sequence version replaced gi:6648222.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L5539
 Center clone name: 390_D_24

 * NOTE: This record contains 175 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 *
 * 1
 * 891: contig of 891 bp in length
 * 892 gap of unknown length
 * 1791: contig of 900 bp in length
 * 1792 gap of unknown length
 * 2706: contig of 915 bp in length
 * 2707 gap of unknown length
 * 3624: contig of 918 bp in length
 * 3625 gap of unknown length
 * 4586: contig of 962 bp in length
 * 4587 gap of unknown length
 * 5534: contig of 948 bp in length
 * 4587 gap of unknown length
 * 5535 gap of 926 bp in length
 * 6461 gap of unknown length
 * 7361 gap of 900 bp in length
 * 8300: contig of 940 bp in length
 * 8301 gap of unknown length
 * 9217: contig of 917 bp in length
 * 9218 gap of unknown length
 * 10157: contig of 940 bp in length
 * 10158 gap of unknown length
 * 11064: contig of 907 bp in length

```
* *
* 11065 11998: gap of unknown length
* *      gap of 934 bp in length
* *      gap of unknown length
* 11999 12918: contig of 920 bp in length
* *      gap of unknown length
* 12919 13842: contig of 924 bp in length
* *      gap of unknown length
* 13843 14750: contig of 908 bp in length
* *      gap of unknown length
* 14751 15703: contig of 953 bp in length
* *      gap of unknown length
* 15704 16635: contig of 932 bp in length
* *      gap of unknown length
* 16636 17572: contig of 937 bp in length
* *      gap of unknown length
* 17573 18510: contig of 938 bp in length
* *      gap of unknown length
* 18511 19442: contig of 932 bp in length
* *      gap of unknown length
* 19443 20365: contig of 923 bp in length
* *      gap of unknown length
* 20366 21284: contig of 919 bp in length
* *      gap of unknown length
* 21285 22179: contig of 895 bp in length
* *      gap of unknown length
* 22180 23144: contig of 965 bp in length
* *      gap of unknown length
* 23145 24060: contig of 916 bp in length
* *      gap of unknown length
* 24061 25030: contig of 970 bp in length
* *      gap of unknown length
* 25031 25949: contig of 919 bp in length
* *      gap of unknown length
* 25950 26914: contig of 965 bp in length
* *      gap of unknown length
* 26915 27855: contig of 941 bp in length
* *      gap of unknown length
* 27856 28764: contig of 909 bp in length
* *      gap of unknown length
* 28765 29706: contig of 942 bp in length
* *      gap of unknown length
* 29707 30657: contig of 951 bp in length
* *      gap of unknown length
* 30658 31593: contig of 936 bp in length
* *      gap of unknown length
* 31594 32536: contig of 943 bp in length
* *      gap of unknown length
* 32537 33459: contig of 923 bp in length
* *      gap of unknown length
* 33460 34397: contig of 938 bp in length
* *      gap of unknown length
* 34398 35299: contig of 902 bp in length
* *      gap of unknown length
* 35300 36229: contig of 930 bp in length
* *      gap of unknown length
* 36230 37172: contig of 943 bp in length
* *      gap of unknown length
* 37173 38122: contig of 950 bp in length
* *      gap of unknown length
* 38123 39048: contig of 926 bp in length
* *      gap of unknown length
* 39049 40013: contig of 965 bp in length
* *      gap of unknown length
* 40014 40927: contig of 914 bp in length
* *      gap of unknown length
* 40928 41851: contig of 924 bp in length
* *      gap of unknown length
* 41852 42766: contig of 915 bp in length
* *      gap of unknown length
* 42767 43716: contig of 950 bp in length
* *      gap of unknown length
* 43717 44632: contig of 916 bp in length
* *      gap of unknown length
```

```
* *
* 44633 45540: contig of 908 bp in length
* *      gap of unknown length
* 45541 46457: contig of 917 bp in length
* *      gap of unknown length
* 46458 47425: contig of 968 bp in length
* *      gap of unknown length
* 47426 48365: contig of 940 bp in length
* *      gap of unknown length
* 48366 49294: contig of 929 bp in length
* *      gap of unknown length
* 49295 50219: contig of 925 bp in length
* *      gap of unknown length
* 50220 51142: contig of 923 bp in length
* *      gap of unknown length
* 51143 52066: contig of 924 bp in length
* *      gap of unknown length
* 52067 53020: contig of 954 bp in length
* *      gap of unknown length
* 53021 53920: contig of 900 bp in length
* *      gap of unknown length
* 53921 54875: contig of 955 bp in length
* *      gap of unknown length
* 54876 55769: contig of 894 bp in length
* *      gap of unknown length
* 55770 56697: contig of 928 bp in length
* *      gap of unknown length
* 56698 57628: contig of 931 bp in length
* *      gap of unknown length
* 57629 58568: contig of 940 bp in length
* *      gap of unknown length
* 58569 59470: contig of 902 bp in length
* *      gap of unknown length
* 59471 60424: contig of 954 bp in length
* *      gap of unknown length
* 60425 61383: contig of 959 bp in length
* *      gap of unknown length
* 61384 62336: contig of 953 bp in length
* *      gap of unknown length
* 62337 63258: contig of 922 bp in length
* *      gap of unknown length
* 63259 64141: contig of 883 bp in length
* *      gap of unknown length
* 64142 65045: contig of 904 bp in length
* *      gap of unknown length
* 65046 66008: contig of 963 bp in length
* *      gap of unknown length

alignment_scores:
Quality: 61.50      Length: 32
Ratio: 2.674      Gaps: 1
Percent Similarity: 71.875      Percent Identity: 37.500

alignment_block:
US-09-331-631-5_COPY_1_32 x AC019039/rev ...

Align seg 1/1 to reverse of: AC019039 from: 1 to: 162168

2 CysMetClnleuGlutHrSerGlyGlnMetArgGArgCysValserGlnC 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
33263 TGCATATCTCTCTCCCTCCCTGCGCAATGTAATAATGTAATCTCCCAATC 33214
18 sAspLysArgPheGlnGlnAspIleAspTrpSer...LysTyrAsp 32
::: :::::|||||:|||||:|||||:|||||:
33213 TGAATGCGCATGGCCACGTGACTTGTGGCCAGTAATAATATAGAG 33168

seq_name: gp_pr4:AC011331

seq_documentation_block:
LOCUS AC011331 175278 bp DNA PRI 06-OCT-1999
DEFINITION Genomic Sequence for Homo sapiens clone H_NH0514011, chromosome 18,
complete sequence.
ACCESSION AC011331
VERSION AC011331.1 GI:6013504
KEYWORDS HTG.
```

```

SOURCE          Homo sapiens
ORGANISM        human.
REFERENCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS         Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE       1 (bases 1 to 175278)
AUTHORS         Huang, E.N., Nascimben, L.U., de la Bastide, M., Vil, D.M.,
                Preston, R., Matero, A., Shah, R., O'Shaughnessy, A.,
                Shekhar, M., Schütz, K., See, L.H., Swaby, I.,
                Dedhia, N.N., and McCormick, R.W.
TITLE           Genomic Sequence for Homo sapiens clone H.NH0514011, Chromosome 18,
                complete sequence
JOURNAL         Unpublished
REFERENCE       2 (bases 1 to 175278)
AUTHORS         Huang, E.N., Nascimben, L.U., de la Bastide, M., Vil, D.M.,
                Spiegel, L.A., Preston, R.R., Matero, A., Shah, R., O'Shaughnessy, A.,
                Rodríguez, M., Shekhar, M., Schütz, K., See, L.H., Swaby, I.,
                Habermann, K., Dedhia, N.N., and McCormick, R.W.
TITLE           Direct Submission
JOURNAL         Submitted (06-OCT-1999) Lita Annenberg Hazen Genome Sequencing
                Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
                Harbor, NY 11724, USA
COMMENT         At location 59145 there is a subclone variation in which one
                subclone calls a 'G' and all the rest call 'A'. From 94100-94174
                there is single-strand/single chemistry. At 148425 there is a
                subclone variation in which one subclone calls 'T' and all others
                call 'G'.
FEATURES        Location/Qualifiers
                source          1..175278
                                /organism="Homo sapiens"
                                /db_xref="taxon:9606"
                                /chromosome="18"
                                /clone="H.NH0514011"
BASE COUNT      46991 a 36801 c 37956 g 53530 t
ORIGIN
alignment_scores:
                Quality:      61.50      Length:      32
                Ratio:        2.674      Gaps:         1
                Percent Similarity: 71.875      Percent Identity: 37.500
Alignment Block:
US-09-331-631-5_COPY_1_32 x ACO11331/rev ..
Align seg 1/1 to reverse of: ACO11331 from: 1 to: 175278
                2 CysMetGlnLeuGlnIuThrSerGlyGlnMetAArgGcysValSerGlnCy 18
                |||:::||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
54662 TGCATATCTCTCTCTCTCTCTCTGCGCATGTGTAATAATGACTTCCCAATC 54613
                18 sasplysArgpneGluGlnUaspIleAspTrpSer...LysTyrAsp 32
                :::::||||:::|||||:::|||||:::|||||:::|||||:::
54612 TGAATGGCATGGCCACGTCGACTTGTCTTGGCCACAGTAATAATATGAG 54567
seq_name: gb_1n1:CELC48E7
seq_documentation_block:
LOCUS          CELC48E7      37701 bp      DNA              INV      23-APR-1997
DEFINITION     Caenorhabditis elegans cosmid C48E7.
ACCESSION      AF000262
VERSION        AF000262.1  GI:1947126
KEYWORDS
SOURCE         Caenorhabditis elegans strain-Bristol N2.
ORGANISM       Caenorhabditis elegans
REFERENCE      Eukaryotes; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
AUTHORS        Rhabditina; Rhabditicoidea; Rhabditidae; Pelodetinae; Caenorhabditis.
REFERENCE      1 (bases 1 to 37701)
AUTHORS        Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M.,
                Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
                Coulson, A., Craxton, M., Dear, S., Du, Z., Dunabin, R., Favello, A.,
                Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
                Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,
                Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,

```

TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans	Nature	368 (6466), 132-38 (1994)		Wamsley, P. and Kramer, J.	The sequence of C. elegans cosmid C48E7	Unpublished (1997)		Waterston, R.	Waterston, R.	Direct Submission	
										Submitted (18-APR-1997)	
										Submitted by:	

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is T0852, 700 bp overlap; 3' cosmid is C10G11, 700 bp overlap. Actual start of this cosmid is at base position 1 of CELC48E7; actual end is at 6598 of CELC10G11

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder(P. Green and L. Hillier, ms in preparation)

FEATURES	source	Location/Qualifiers
		1..37701
		/organism="Caenorhabditis elegans"
		/strain="Bristol N2"
		/db_xref="taxon:6239"
		/chromosome="I"
		/clone="C48E7"
gene		121..624
		/gene="C48E7.7"
		join(121..200,246..624)
CDS		/gene="C48E7.7"
		/codon_start=1
		/evidence=not_experimental
		/protein_id="AB52932.1"
		/db_xref="GI:1947137"
		/translation="MOSINILRFAMLLIPIYNGDPTAAVATIKTEDEYALITTYAKVAATEPEPTPEVQTTEAVEETTEAVVEAQETDAETNAVDTPAANNIEATTEGASPPSSSTVPAVPTSVDDLGETSTNAKRAYNCTFEVPMVVALIO"
gene		4876..14780
		/gene="C48E7.6"
		join(4876..4956,5006..5146,5384..5730,5773..5907,6158..6368,6463..6861,6935..7012,7062..7379,7606..7829,7925..8050,8102..8334,8566..8755,8802..8980,9025..9205,9256..9411,9456..9535,9715..9783,9857..9991,10048..10611)
CDS		

```

CBS      join(18420..18517,18563..18650,18698..18808,19037..19110,19210..19347,19393..19642,19687..20002,20050..20176)
        /gene="C48E7.3"
        /note="similar to CCAAT/enhancer-binding protein"
        /codon_start=1
        /evidence=not_experimental
        /protein_id="AB52934.1"
        /db_xref="GI:1947129"

gene     join(20378..20440,20486..20578,20641..20762,20811..21099,21566..21885,21945..22089,22152..22363,22461..22856)
        /gene="C48E7.2"
        /note="region of weak similarity to a short region of DNA-directed RNA polymerase III, 74 kd subunit"
        /codon_start=1
        /evidence=not_experimental
        /protein_id="AB52935.1"
        /db_xref="GI:1947130"
        /translation="MGGGKYEYOMCVIIEIDIFGRIVAKVMEITLKESCOLSTLYPKLKGDISNLTIRKSYSVLINFGVFTHLDNTNNRTTYADSSQIOHVISSPACILATLYGPSAEISCEVISOGRLVYSETRIQARDENFAVSVDIKSFEEELAAAFILRPVSVESEHGCPQFVTNPDPRAISTNIMEKEEGTSDTSNNKKRADSPDEGDAGQYRIWNIRENGYLDELIIIDYLIIGSGSKGMTNLANKKEETIEDPGDRQTITOVANLMFKINERTAPKPMSESIIVASLTDLRQLKENGILANPADIELAQIIVDESdGiIRKGE SNGGITVIDIRAIRQICRHCESLIREOFEGEARVRMLLENHRLHEDEQVEKLSMSGEAREEMIALYALVEEGYFNKPVGRSNDNQPSFTFLFYVDVPTIRGLVEYTKLVRNLIVROHRHTREKNSVLDKANVOPIENIRSSPOEDSKIAQIAEVEWMLTPGDR ACQNRPRKOATILAAQDSSIVLSFKFLASGEN"
        25399..29605

alignment_scores:
    quality: 61.00          length: 28
    ratio:   2.542         gaps: 0
    percent similarity: 85.714    percent identity: 28.571

alignment_block:
US-09-331-631-5_COPY_1_32 x CELC48E7/rev ..

Align seg 1/1 to reverse of: CELC48E7 from: 1 to: 37701

      4 glnleuglurhtSerGlyImMetArgARGyValSerGInCysAspLy 20
      ::::||||| ||||: :::|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
36576 AAATCGAACAAGTAGGCGTAACCCGAAGAAGTGTCGCAACATGTCCGAGAA 36527

      20 sArPhheGlunLaspIleaSpITpSerLvsTyR 31
      :|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
36526 AAGGCTGAAGCAAATGTGCGATTCTGTATTAT 36493

seq_name: gb_pl2:AC003033

seq.docmentation_block:
LOCUS   AC003033              84254 bp            DNA             17-DEC-1999
DEFINITION Arabidopsis thaliana chromosome II section 184 of 255 of the complete sequence.
ACCESSION AC003033 AE002093
VERSION    AC003033.2 GI:6598385
KEYWORDS
SOURCE    thale cress.
ORGANISM  Arabidopsis thaliana
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eudiphyloptes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
REFERENCE 1 (bases 1 to 84254)

```

AUTHORS Lin, X., Kaul, S., Rounsley, S.D., Shea, T.P., Benito, M.-I., Town, C.D., Fujii, C.Y., Mason, T.M., Bowman, C.L., Barnstead, M.E., Feldhahn, T.V., Buell, C.R., Ketchum, K.A., Lee, J.J., Rensing, C.M., Koo, H., Moffat, K.S., Cronin, L.A., Shen, M., Vanaken, S.E., Umayam, L., Tallon, L.J., Gill, J.E., Adams, M.D., Carreira, A.J., Creasy, T.H., Goodman, H.M., Somerville, C.R., Cohenver, G.P., Preuss, D., Nierman, W.C., White, O., Eisen, J.A., Salzberg, S.L., Fraser, C.M. and Venter, J.C.

TITLE Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*

JOURNAL Nature 402 (6763), 761-768 (1999)

MEDLINE 20083487

REFERENCE 2 (bases 1 to 84254)

AUTHORS Lin, X.

JOURNAL Direct Submission

COMMENT Submitted (13-DEC-1999) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

On Dec 17, 1999 this sequence version replaced gi:2702261. The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (<http://www.tigr.org/db/at/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (<http://artur.ebm.ornl.gov/pub/xgrail/>), GeneFinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://genomic.stanford.edu/GENSCAN.html>), and NetPlantGene (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to the bottom of the chromosome.

We thank the CSHL/MashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13J16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: atltigr.org.

FEATURES

source 1.84254 Location/Qualifiers

organism="Arabidopsis thaliana"

cultivar="Columbia"

db_xref="taxon:3702"

chromosome="2"

complement(1..1230)

note="Sequence from clone F24L7"

repeat_region 297..343

family="GAA)n"

331..>1332

gene="At2g32820"

gene="At2g32820"

note="F24L7.4: predicted by genefinder and genscan"

331..1332

CDS

gene="At2g32820"

note="hypothetical protein"

/codon_start=1

protein_id="BAB91965.2"

db_xref="GI:6598386"

translation="WMKTAITERTDLPRLVSTAIRASNSQNSADVERCPDVLRLKGLTNSKNSHSRVLPLESLDHNPKIRTEAHVLFWMKTFYSSGQNSSTCNKPNLKTKVYKACSELKKNNEQLSHGFVNLKAKETVFGMKKNEDERSRVHETEMKOGDSKSFALMTIEKKQASPIQSGSKNPSRSGAGETGYIKLIRKPDCCSVSRPLMKKHQDVKAEPENKTCIVYKKSPEMLFEFAKKSADVANAAGFLAAKEASICVDILLMKFSIISTAIETERRIMEKLERITKDKKICNALALLHHNRQTRNQQ"

complement(1231..81848)

note="Sequence from clone T21L14"

complement(join(<2300..3151,3250..3722,3900..>4203))

gene="At2g32830"

complement(<2300..>4203)

gene="At2g32830"

note="T21L14.23"

complement(join(2300..3151,3250..3722,3900..4203))

gene="At2g32830"

codon_start=1

product="putative phosphate transporter"

protein_id="BAB91982.1"

db_xref="GI:2702279"

translation="MAKKGKVELNALDAKTOYHFTAVIAGMGFTDAYDLFISLVRKLGRIYHYDSSKRGTLPPNVAANGVAFGTLAQLFPGMLGDKLRKRYGITLMLAVICSGISGSPGANGVMAITFGFPGIGDGPISLATIMSPYAKRRAFLVAAPAMGPGIILAGIYSLVSSFDFAKAPYENDVPGSTVDADYVNRVLMFGAIPALITTYWPKMKETARKYALVALARKTKQASDKSKVLOYDLAEBAQNSNSNPNPTTGLETFRRARRGHLLGTTTTLFLDLAYSSNLFQDITYALGIMPAETMAHAHEVTVSKAQTALACGTGVTGVTFAFDILSRFFLOMGFTMTJEMFALAIPYDHRHRENRIGILIMYSLETFPANGPNATFTVPAEIPPARLRSHCHISAAGKGAIVGAFGEFLYAAOSSSEKTDGPGICVRSLMLACVNLGITYFTLVPESKCKSLDEISREDESGDPTVEMTVANGRKRPV"

5483..5600

family="TAA)n"

complement(join(<7648..7741,7838..7917,8011..8118,8228..8305,8403..8442,8533..8585,8715..8747,8829)..>9383))

gene="At2g32840"

complement(<7648..>9383)

gene="At2g32840"

note="T21L14.22: predicted by genefinder"

complement(join(7648..7741,7838..7917,8011..8118,8228..8305,8403..8442,8533..8585,8715..8747,8829)..9383))

gene="At2g32840"

note="hypothetical protein"

codon_start=1

protein_id="BAB91981.1"

db_xref="GI:2702278"

translation="ASTDPNPNPMPNPSVSAVSTPIYTAESPVTQPNVITPSSQPOPQPASSYRAIARLHRHNPONITPTPLPRRSNTVNSHOPHPDPSLIYFPGSSGRGFPTRPVARNNSVADPYGSGSPGCTPRGCVYGHQFVSNLDPNKOFRRAHPQNOOSPOLGSGHMKGVPHLPQRLVASEVVEATPSPSILDSNGHKAARSDDALVIVAKRRYRTTEGASLSLCSRNGAHEGICRIDMCTCLPRLPDKYETSLPKDLVEPAICEDKEDESEYKHLSESDLLKRIIDRAKVARLREERIKRIYAKKALALLPPEEOCRNE"

complement(join(9789..11142,11327..11576,11666..11716,11802..11967,12058..>12524))

gene="At2g32850"

complement(9789..>12524)

gene="At2g32850"

note="T21L14.21"

complement(join(9895..11142,11327..11576,11666..11716,11802..11967,12058..12524))

gene="At2g32850"

codon_start=1

product="putative serine/threonine protein kinase"

protein_id="BAB91980.1"

db_xref="GI:2702277"

translation="WMKTFPAQKPEAPLDEGRFLEIGLKQYVRNVIAEGGESSYILAQVNHASROYALKHMTICNDEESLELVMEISVLSKSLKHPVNTLYAGHILDMGNKKBALMAFCGSLVDLENRGAGYEKQALITFEDVQNAVFAFMGOSPRALHARDLKAENLILSSDGOMKLCDFGSGVSTNHKIFERRAEEMGIEEDINIRKTYTPTYPAPEMWDLFRR

EMISEKDIWALICGLERICYENFMADGESKMLIOLGNRRIESPSPYSEFIINDLIKEM
LOASPERDITOWKSOFLIIVSSOLRCKMKSITLSTSSKATKPSAPRSPRPPPPS
SGEDSGCPGATFATWOTHAFTSVSEDNKMEFEEDNPSTSKSEKVRADSHQPKPS
PVREALGIQRNDKDDLETTISQNTTPPAANMFKVRKDAFNSFVADFOTIKLASKA
PGKEALEAEIQRLEKTKQKSEKAEITAKPEKLSAICRSORVOLODKOTLAKSA
SPSPRSNONOPSPGMHSMSTPSVLETLVLSLSDTFTKYEINYLLIMFEIKOR
DKMGDTWELSOEDRSMSTGSDTNSOPDARKVPMASASGNNTINOSVTRRSK
PASACGTGCFEGPMEFTESFRAATSAATASAGTGRSMGSGNSTQRGNCKMKNQ
KTPACMGAGP"

gene
join(<14780, 15118, 15202, 15271, 15365, 15423, 15533, 15608,
15742, 15819, 15934, 16021, 16139, 16398, 16541, 16653,
16793, 17010, 17342, 17479, 17581, 17689, 17849, >18143)
/gene="At2g32860"
<14780, >18143
/gene="At2g32860"
/note="T21L14.20"
join(14780, 15118, 15202, 15271, 15365, 15423, 15533, 15608,
15742, 15819, 15934, 16021, 16139, 16398, 16541, 16653,
16793, 17010, 17342, 17479, 17581, 17689, 17849, 18143)
/gene="At2g32860"
/codon.start=1
/product="putative beta-glucosidase"
/protein_id="FAB91979.1"
/db_xref="GI:2702276"
/translation="MAVATLTFLGLLATLTSTLSFNADARPOSDDELGTITGHQOT
SFDELDIVIGPHTAVDDEDIDMDKGTGVPQNLNDDDLGTIGEEFIHQDPAD
FETGTSAYAGVEGAKKSGRGLTSMDETFHMEPEVKQNGDEVDYTRYKDDIK
LMELNTNGRFSISWTRILPYGTITKGVNEBVKRYNLLIMDLANGIOPVTLFHW
ESPALCEKMGQFLNRIYVDEREPANFOCKGPDGRKMAATPNEFSYAVAGYKCK
KAGRCCKNOAPKPCPTGDSSEEPYIAHNDIILAHLAANDVEFNCKYBEGGKITGLV
SHMEPRKNSSEDEVKAKRSLEYQLGFEURLPULTGQYPAEMLEVDNITLRETFEES
EKLRKSLDEFLVNLGAFESTPLAKVNSSQLNENIDLKRWVETLITNNLSLPLQOTSM
GIAYIPAGLNIILKHKIKDEYMDPEIYIMENGMDEIYGTNRITTEATINDYGRREFKSH
ILIMGKSLRMDKVRLLKGYIWSLMDNPEMDGVKAFVFGIYVDYDNMKNRYTRSSGNW
LSEEDSKETFLKCYEFGHEKGYAKLPDVELEPENSQLSYRSDFM"

misc_feature
/note="molecular marker 96191"
complement(join(19158, 19966, 20113, 20413, 20521, 20672,
20766, 20895, 20993, 21068))
/gene="At2g32870"
complement(19158, 21068)

gene
alignment_scores:
Quality: 61.00 Length: 30
Ratio: 3.389 Gaps: 0
Percent Similarity: 60.000 Percent Identity: 36.667

alignment_block:
US-09-331-631-5_COPY_1_32 x AC003033 ..
Align seg 1/1 to: AC003033 from: 1 to: 84254

2 CysMetGlnLeuGlnIhrSerGlyImeTarGArgCysValIserGlnCys 18
|||||:|||||: ||| ||||| ||
9237 TCTTTGAGGTGTGAGCTGAGACGCGGCTATGACTGTGTAGGTTGTG 9286
|||||:|||||: ||| ||||| ||
18 sasplysargpneglugluaspiliasptirserlystyr 31
|:|||||: ||| ||||| ||
9287 TCACGCGAGATTCCATCGGTGACGATTGAGTGCATGAC 9326
|:|||||: ||| ||||| ||